Genome wide association study on beef production traits in Marchigiana cattle breed

Sorbolini S.,Bongiorni S.,Cellesi M.,Gaspa G.,Dimauro C.,Valentini A. and Macciotta N,
Genome wide association study on beef production traits in Marchigiana cattle breed,
J. Anim. Breed. Genet., 134, 2017, 43-48,
doi:10.1111/jbg.12227

Genome Wide Association Study on beef production Traits in Marchigiana Cattle breed

Silvia Sorbolini[1], Silvia Bongiorni[2], Massimo Cellesi[1], Giustino Gaspa[1], Corrado Dimauro[1], Alessio Valentini, and Nicolò Pietro Paolo Macciotta[1*].

[1] Dipartimento di Agraria, Sezione di Scienze Zootecniche, Università degli Studi di Sassari, 07100, Sassari, Italy;

[2] Dipartimento per l'Innovazione dei Sistemi Biologici Agroalimentari e Forestali DIBAF, Università della Tuscia, Viterbo, Italy.

*corresponding author: Nicolò Pietro Paolo Macciotta, Dipartimento di Agraria, Sezione di Scienze Zootecniche Università degli Studi di Sassari, 07100, Sassari, Italy. E-mail: macciott@uniss.it

Summary

A genome-wide association study was carried out on a sample of Marchigiana breed cattle to detect markers significantly associated with carcass and meat traits. Four hundred and nine young bulls from 117 commercial herds were genotyped by Illumina 50K BeadChip assay. Eight growth and carcass traits (average daily gain, carcass weight, dressing percentage, body weight, skin weight, shank circumference, head weight, carcass conformation) and two meat quality traits (pH at slaughter and pH 24 hours after slaughter) were measured. Data were analyzed with a linear mixed model that included fixed effects of herd, slaughter date, fixed covariables of age at slaughter and SNP genotype, and random effects of herd and of animal. A permutation test was performed to correct SNP genotype significance level for multiple testing. A total of 96 SNPs were significantly associated at genome-wide level with one or more of the considered traits. Gene search was performed on genomic regions identified on the basis of significant SNP position and level of linkage disequilibrium. Interesting loci affecting lipid metabolism (*SOAT1*), bone (*BMP4*) and muscle (*MYOF)* biology were highlighted. These results may be useful to better understand the genetic architecture of growth and body composition in cattle.

Keywords: SNP chip, GWAS, bovine, productive traits

Introduction

The recent availability of high throughput SNP platforms for several livestock species has revitalized the search for DNA markers associated to phenotypic variation in complex traits of economic importance (Bush and Moore 2012). Genome-wide association studies (GWAS) represent a first step toward the understanding of molecular and cellular mechanisms underlying phenotypic expression of complex traits (Jiang *et al.* 2010; Korte and Farlow 2013).

48 Genomic approaches are expected to have a great impact on traits that are difficult and expensive to
49 measure. An example are *post-mortem* traits in beef cattle. Dressing percentage, carcass
50 composition, and meat quality are difficult to obtain and relate to animals retained for selection.
51 Recent GWAS studies have detected associations between SNPs and beef traits, suggesting
52 *myostatin*, *DGAT1* and *leptin receptor* as candidate genes (Jiang *et al*. 2010).
53 Local beef breeds are important for typical production systems and for crossbreeding with
54 specialized breeds. GWAS carried out on local breeds may provide useful insights in the genetic
55 determinism of meat traits by picking up genetic variation no longer detectable in cosmopolitan
56 breeds. In Italy there are several local beef cattle breeds. They differ in selection history, trait
57 phenotypic expression, and genetic background (Sorbolini *et al.*, 2015). The Marchigiana breed is a
58 typical example. It originated from the Asiatic long-horned (*Bos primigenius*) cattle and moved to
59 Italy from Central Asian steppes during invasions in the sixth/seventh century C.E. (Trombetta *et*
60 *al.* 2005). Beef traits were improved by crosses with Chianina and Romagnola cattle in the second
61 half of the nineteenth century. The current Marchigiana is the result of a breeding program started
62 after the above mentioned cross-breeding. At present, it is the second beef breed of Italy with about
63 52,344 hd registered in the Herdbook. It is characterized by a strong adaptability to harsh
64 environmental conditions, great precocity, fertility and a remarkable aptitude for meat production
65 (Balasini 1981) due to well-pronounced muscle development and fine bone structure and skin. For
66 these reasons it has also been exported to countries such as United States, Canada, Brazil,
67 Argentina and Australia ("http://www.anabic.it/")
68 In the present work, a GWAS was carried out on a sample of 409 Marchigiana young bulls farmed
69 in commercial herds, genotyped with the Illumina Bovine SNP50 BeadChip. The study was aimed
70 at identifying chromosome regions harbouring new putative candidate genes affecting meat and
71 carcass quality traits in beef cattle.
72

Material and Methods

Animals and phenotypic data

Four hundred and nine Marchigiana young bulls from 117 commercial herds were slaughtered between 16 and 24 months of age. Phenotypes of ten different growth, carcass and meat quality traits were recorded at the slaughter house: body weight (BW), average daily gain (ADG), carcass weight (CW), dressing percentage (DP), skin weight (SW), shank circumference (SC), head weight (HW), carcass conformation according to the European grid based on muscularity and fat content (SEUROP) evaluation system (CC), pH at slaughter (pH) and pH 24 hours after slaughter (pH24h). pH at slaughter and 24h after slaughter were measured on the *longissimus dorsii* muscle with the HI 99 163 pHmeter (Hanna instruments).

Genotypic data

Genomic DNA was extracted from whole blood samples gathered immediately before slaughter using the NucleoSpin 96 Blood Kit (Macherey-Nagel) according to manufacturer's instructions. All 409 animals were genotyped using the Illumina 50K BeadChip assay. SNP editing was on call rate (>99%) and minor allele frequency (>1%). Animals having more than 2,5% of missing genotypes were discarded. A total of 43,313 markers were retained after edits.

Statistical Analysis

Data were analyzed using the following mixed linear model:

$$Y = D + bAGE + bSNP + a + h + e \quad [1]$$

where:

Y = record for the the considered trait;

D = fixed effect of slaughter date (46 levels);

bAGE = fixed covariable of age at slaughter in months ;

bSNP = fixed covariable of SNP genotype (coded as 0, 1, 2 according to the number of second allele)

a =random additive genetic effect of the animal.

h = random effect of the herd (114 levels);

e = random residual.

The animal effect was assumed to be normally distributed ~$N(0,\mathbf{G}\sigma^2_a)$ where **G** is the genomic relationship matrix and σ^2_a is the additive genetic variance. **G** was calculated according to VanRaden (2008) as:

$$\mathbf{G} = \frac{\mathbf{ZZ}'}{2\sum p_i(1-p_i)}$$

where **Z** is the matrix of individual genotypes scaled by allele frequencies (p_i) expressed as differences from 0.5.

A modified version of the experimentwise empirical threshold proposed by Churchill and Doerge (1994) was used to correct SNP statistical significance for multiple testing. In a first step, single marker analysis was performed with model [1]. Significant markers ($P<0.01$) were retained. In the second step, 10,000 permutations were performed for each significant marker by shuffling SNPs across animals, while keeping invariant the other factors included in model [1] (Anderson and Ter Braak 2003). The bottom 5% of α probabilities of test statistics for each marker (SNP_ALPHA) were retained. Then SNP_ALPHA for all SNPs were put in the same column, and the 5th percentile was kept as a critical threshold for declaring significant at $P<0.05$ tests performed in the first step. Statistical analyses were performed using SAS 9.2 (SAS/STAT software version 9.2, SAS Institute, Inc. Cary, NC, USA).

Putative candidate genes identification

Gene search was performed on chromosome regions defined by positions of significant SNPs according to the sixth draft of bovine genome assembly (UMD3.1/bosTau 6) UCSC Genome Browser Gateway (http://genome.ucsc.edu./). Windows of variable amplitude in Mb were defined based on linkage disequilibrium of the specific genomic region (Macciotta *et al.* 2015). For each significant SNP the squared coefficient (r^2) statistic with all other SNPs positioned in the same chromosome was calculated (Table S1). Distance between the significant SNP and the furthest SNP having an $r^2 > 0.10$ was calculated and added upstream and downstream to the position of significant marker. SNP not in LD with other markers were not considered for gene discovery. Finally, specific functional analysis and biological roles of annotated genes were investigated by an accurate literature search and databases consultation such as GeneCards (www.genecards.org), National Centre for Biotechnology Information (www.ncbi.nlm.nih.gov), Proteinatlas (www.proteinatlas.org). Gene names and symbols were derived from HUGO Gene nomenclature database (www.genenames.org).

Results

Significant SNPs and association analyses

A total of 96 SNPs were found to be associated with seven out of ten considered traits (ADG, CW, DP, BW, HW, SC and pH) (Table S1). As an example, figure 1 reports the Manhattan plot for ADG.

No significant SNP were found for pH24h, SW and CC. The largest number of significant markers associated with different traits was found on BTA2 (14 SNPs), followed by BTA6 (11 SNPs) and 8 (10 SNPs). Chromosomes 9, 11, 12, 18, 19, 23, and 29 showed only one significant marker. BTAs 13, and 27 did not show any associated marker.

Significant markers of BTA2 were associated with five different traits (ADG, CW, DP, pH, and SC) followed by BTA8 with four (ADG, HW, SC, pH) and 26 with three (HW, SC, pH) . Finally, a

146 total of two SNPs resulted associated with two traits (ADG and BW); rs43272238 on BTA1, and
147 rs41662409 on BTA16.
148
149 Average Daily Gain

150 Forty-five significant markers were detected for ADG. Chromosome 6 showed the highest number
151 of SNP associated with this trait (10). BTAs 5, 15, 22, 24 and 28 contained only one significant
152 marker associated with ADG. A SNP located on BTA10 between 65,7 and 67,5 Mb (*rs41568676*)
153 flagged a region where the *bone morphogenetic protein 4* (*BMP4*) gene maps (Table1). On BTA14
154 the rs41631408 at 57469150 bp pointed out the *thyrotropin-releasing hormone receptor* (*THRH*)
155 locus. Other significant markers associated with ADG identified several distinct genes involved
156 primarily in cellular processes such as growth and proliferation (*IFRD1,CGRRF1, TGFB2*), but
157 also genes involved in general metabolic pathways such as (*SPTLC1*, *UTG1A6* and *UTG1A1*) or
158 specific pathways such as carbohydrate metabolism (*ALDOA*) and lipid metabolism (*SOAT1*)
159 (Table1).

160
161 Shank circumference
162 Table S1 reports the13 significant markers found to be associated with SC. After ADG, it was the
163 trait with the highest number of significant associated markers. Three of them were found on
164 BTA14 and two on BTA8. However, no annotated genes were retrieved in the corresponding
165 chromosomic regions.
166
167 Dressing Percentage
168 Twelve SNPs were found significantly associated with DP. Eight out of 12 were located in a large
169 chromosomic region between 1,0-5,2 Mb on BTA2. These SNPs were in close proximity with a

QTL that contains the *myostatin* (*MSTN)* locus and two other genes that have a role in muscle biology (*SLC40A1* and *COL5A2.)* On BTA9 at 288595 bp from the significant SNP *rs 41662464* map the *connective tissue growth factor* (*CTGF*), a gene involved in chondrocyte proliferation.

Carcass Weight

In this study, 9 significant markers distributed over seven different autosomes were associated with CW. Four SNPs were found on BTA5 (Table S1). On BTA2 the SNP *rs109168082* at 129,8 Mb tagged to the *PNRC2*, *GALE* genes (Table 1). On BTA23 the validated mRNA sequence of *ATP-binding cassette, subfamily F (GCN20), member 1* (*ABCF1*) is annotated close to the *rs110277462* marker.

Head Weight

A total of 7 significant SNPs were associated with HW. BTA7 harbored the largest number of markers (n = 2) associated with this trait (Table S1) whereas BTAs 5,11,16,20 and 26 showed a single significant marker.

Body weight

A total of 5 significant SNPs were found associated with BW (Table S1). Few annotated genes were retrieved in the intervals surrounding these SNPs. Three significant markers were shared with other traits examined in this study. On BTA7, a significant marker (*rs42691441*) associated to the BW and located at 68,070,311 bp was also associated with HW. The annotated sequence nearest the marker was the *CCR4-NOT transcription complex, subunit 8 mRNA (CNOT8).*

pH at slaughter

Five significant markers were found to be associated with pH at slaughter (Table S1). A single associated SNP was on BTAs 2, 3, 14, 17, and 26. No suggestive genes were found for this trait.

Discussion

Growth performance and growth-related traits such as body size and weight or average daily gain, have a crucial role in livestock due to their influence on meat production. Average daily weight gain is one of the most important traits for assessment of animal growth and it is a component of most economic indices. In livestock, discovering and understanding genes and molecular mechanisms underlying differences in ADG could clarify relationships among weight gain and other important traits such as body composition or feed intake (Santana *et al.* 2014).

Marchigiana cattle have been selected for meat production (a trait with a medium to high heritability) over the last twenty years. Aim of this study was to identify candidate genes associated with beef production traits in this breed. The total number of significant associations detected in this GWAS was in general agreement with literature (Snelling *et al.* 2010; Rolf *et al*. 2012).

SNPs significantly associated to ADG flagged regions where genes involved in the metabolism of sugars and lipids are located. This is in general agreement with cattle physiology because these metabolic pathways may have a significant influence on average daily gain. An interesting outcome of the present study is represented by the association between ADG and two markers (*rs41662409* and *rs110397182)* located on BTA 16. These associations underline *Sterol-O-Acyltransferase 1* (*SOAT1*) and *transforming growth factor, beta 2* (*TGFB2*) genes, respectively. In particular *SOAT1* was already reported as a candidate gene in beef cattle (Jiang *at al.* 2009). *SOAT1* encodes for an enzyme that is involved in steroidogenesis and lipogenesis/lipolysis network.

Another promising candidate gene for ADG was *TGFB2.* This gene regulates cell proliferation and differentiation and it was already reported as a locus involved in extracellular matrix organization of muscle development (Guo *et al.* 2015). Moreover, polymorphisms at *TGFB2* were associated

220 with growth traits in chicken (Mojtaba *et al.* 2013). A significant marker (*rs41631408*) located on
221 BTA14 between 57,4-57,5 Mb highlighted the *thyrotropin-releasing hormone receptor* (*TRHR*).
222 This gene encodes for the receptor responsible of thyrotropin hormone (TRH) release. In mammals
223 *THR* is involved in somatotropin (GH) secretion, regulation and activity (Harvey 1990). The
224 relationship between blood concentration of GH and growth has long been known and positive
225 effects of THRH on growth and carcass characteristics in beef cattle performances were already
226 reported by Enright et al. (1993).
227 Finally, the marker *rs43395215* found to be associated associated with ADG tagged a putative
228 candidate gene, *interferon-related development regulator* (*IFRD1*), involved in adipocyte
229 proliferation, growth and differentiation.
230 Carcass weight and dressing percentage represent economically important traits for livestock
231 production. However, in recent years, meat quality has also received more attention as economically
232 important. Phenotypic traits such as tenderness, marbling and unsaturated fat content are
233 considered essential in the beef industry. Dressing percentage trait is an estimate of amount of
234 saleable product derived from a given carcass (Casas *et al*. 2003). The *MSTN* locus, encoding
235 myostatin, is one of the most studied genes in beef cattle (Djiari *et al.* 2013). Polymorphism at this
236 single autosomal locus causes double muscle phenotype. Several mutations have been previously
237 reported in many cattle breeds for *MSTN* (Djiari *et al.* 2013). In mammals, polymorphisms in this
238 locus result in muscle hyperplasia caused by inactivation of the negative regulator of myogenesis
239 (McPherron and Lee 1997). *MSTN* mutations are associated with increased muscle mass, carcass
240 yield, meat tenderness and a reduction of collagen content in cattle (Esmailizadeh *et al*. 2008).
241 Besides economic benefits, double muscled phenotype implies undesirable consequences such as
242 reduced fertility, low calf viability and dystocia (Bellinge *et al*. 2005). A point mutation consisting
243 of a G/T transversion in the third exon of *MSTN* has been reported in Marchigiana (Marchitelli *et*
244 *al*. 2003). This variant has a rather low frequency in the population, probably due to the careful

breeding policy of breeders that want to avoid negative effects on reproduction. However extreme double-muscling individuals are still observed (Marchitelli *et al*. 2003). Also SNP in the promoter region of this gene may influence muscularity and therefore DP (Crisà *et al.* 2003). Significant markers found in this study identify a QTL region where *MSTN* and other neighboring genes such as *Collagen, typeV, alpha 2* (*COL5A2*) and *Solute carrier family 40, member A1* (*SLC40A1*) involved in muscle biology and collagen biosynthesis were located. This result is in agreement with previous reports for beef cattle *(*Pintus *et al*. 2014, Saatchi *et al*. 2014).

Acknowledgements

This work was supported by the Italian Ministry of Agriculture (grants SELMOL and INNOVAGEN).

References

1) Anderson M.J., Ter Braak C.J.F. (2003) Permutation tests for multi-factorial analysis of variance. J. Stat. Comp. Simul., **73**, 85-113.

2) Balasini D. (1981) Razze bovine da carne. Incroci e loro prospettive. Edagricole.

3) Bellinge R.H.S., Liberles D.A., Iaschi S.P.A., O'Brian P.A., Tay G.K. (2005) Myostatin and its implications on animal breeding: a review. Anim. Genet., **36**, 1-6.

4) Bush W.S., Moore J.H. (2012) Chapter 11: Genome-wide association studies. PLoS Comput. Biol., **8**, e1002822.

5) Casas E., Shakelford S.D., Keele J.W., Koohmaraie M., Smith T.P.L., Stone R.T. (2003) Detection of Quantitative trait loci for growth and carcass composition in cattle. J Anim. Sci., 81, 2976-2983.

6) Churchill G.A., Doerge D.R. (1994). Empirical Threshold Values for Quantitative Trait Mapping. Genetics., **138**, 963-971.

7) Crisà A., Marchitelli C., Savarese M.C., Valentini A. (2003) Sequence analysis of myostatin promoter in cattle. Cytogenet.Genome Res., **102,**. 48-52.

8) Djari A., Esquerre D., Weiss B., Martins F., Meersseman C., Boussaha M., Klopp C., Rocha D. (2013) Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. BMC Genomics., **14**, 307.

9) Enright W.J., Prendiville D.J., Spicer L.J., Striker P.R., Moloney D.P., Mowles T.F., Campbell R.M. (1993) Effects of growth hormone-releasing factor and (or) thyrotropin-releasing hormone on growth, feed efficiency, carcass characteristics, and blood hormones and metabolites on beef heifers. J. Anim Sci., **71**, 2395-2405.

10) Esmailizadeh A.K., Bottema C.D., Selick G.S., Verbyla A.P., Morris C.A., Cullen N.G., Pitchford W.S. (2008) Effects of the myostatin F94L substitution on beef traits. J. Anim. Sci., **86**, 1038-1046.

11) Guo B., Greenwood P.L., Cafe L.M., Zhou G., Zhang W., Dalrymple B.P. (2015) Transcriptome analysis of cattle muscle identifies potential markers for skeletal muscle growth rate and major cell types. BMC Genomics., **16,** 177.

12) Harvey S. (1990) thyrothropin-releasing hormone: a growth-hormone releasing factor.J. Endocrinology, **125**, 345-358.

13) Jiang Z., Michal J.J., Chen J., Daniels T.F., Kunej T., Garcia M.D., Gaskin C.T., Busboom J.R., Alexander L.J., Wright Jr R.W., MacNeil M.D. (2009) Discovery of novel genetic networks associated with 19 economically important traits in beef cattle. Int. J. Biol. Sci., **5,** 528-542.

13) Macciotta N.P.P., Gaspa G., Bomba L., Vicario D., Dimauro C., Cellesi M., Ajmone- Marsan P. (2015) Genome-wide association analysis in Italian Simmental cows for lactation curve traits using a low-density (7K) SNP panel. J. Dairy Sci., **98**, 1-11.

14) Marchitelli C., Savarese M.C., Crisà A., Nardone A., Ajmone-Marsan P., Valentini A. (2003) Double muscling in Marchigiana beef breed is caused by a stop codon in the third exon of *myostatin* gene Mamm Genome., **14,** 392-5.

15) Mc Pherron A.C., Lee A. C. (1997) Double-muscling in cattle due to mutations in the myostatin gene. Proc. Natl. Acad Sci., USA. 94, 12457-12461.

16) Pintus E., Sorbolini S., Albera A., Gaspa G., Dimauro C., Steri R., Macciotta N.P.P. (2014) Use of locally weighted scatterplot smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds. Anim. Genet., **45**, 1-11.

17) Rolf M.M, Taylor J.F., Schnabel R.D., McKay S.D., McClure M.C., Nborthcult S.L., Kerley M.S.,Weaber R.L. (2012) Genome-wide association analysis for feed efficiency in Angus cattle. Anim Genet., **43**, 367-374.

18) Saatchi M., Schnabel R.D, Taylor J.F., Garrick D. J. (2014) Large-effect pleiotropic or closely linked QTLsegregate within and across ten US cattle breeds. BMC Genomics, **15**, 442-458.

19) Santana M.H.A., Utsunomya Y.T., Neves H.H.R., Gomes R.C., Gatcia J.F., Fukumasu H., Silva S.L., Leme P.R., Coutinho L.L., Eler J.P., Ferraz J.B.S. (2014) Genome-wide association study for feedlot average daily gain in Nellore cattle (*Bos indicus*),.J. Anim. Breed. Genet. **131**, 210-216.

20) Snelling W.M., Allan M.F., Keele J.W., Khuen L.A., McDaneld T., Smith T.P.L., Sonstegard T.S., Thallman R.M., Bennett G.L. (2010) Genome-wide association study of growth in crossbred beef cattle. J. Anim. Sci., **88**, 837-848.

21) Sorbolini S., Marras G., Gaspa G., Dimauro C., Cellesi M., Valentini A., Macciotta N.P.P (2015) Detection of selection signatures in Piemontese and Marchigiana cattle, two breeds with similar production aptitudes but different selection histories GSE., **47**, 52-65.

22) Trombetta M.F., Mattii S., Sbarra F., Palazzo R., Caimmi D., Falaschini a., Forabosco F., Filippini F. (2005) Observations at slaughtering time: carcass characteristics of Marchigiana beef. Preliminary results. Proceedings of the 4th World Italian Beef Cattle Congress, Gubbio (Italy) April 29th- May 1st 549-554.

24) VanRaden P. (2008) Efficient Methods to Compute Genomic Predictions. J. Dairy Sci., **91**, 4414–4423.

Supporting Information

Table S1: List of significant markers associated with the traits under study.

Tables

Table 1.

Figures

Figure 1.

Captions for Tables

Table 1: Putative candidate genes associated with *in vivo* and *post mortem* phenotypes under study.

Captions for Figures

Figure 1. Genome-wide association study of average daily gain. The dashed line corresponds to a permutation treshold of 0.05.